Corpus oraux : Essai de segmentation automatique

Noura Tigziri
Département de langue et culture amazighes
Université Mouloud Mammeri de Tizi-Ouzou

Nora.tigziri@gmail.com

Introduction : Présentation du projet

Notre projet consiste en la mise en place d'une banque de données de corpus oraux, numérisés, transcrits et annotés pour la langue amazighe qui soit exploitable à des fins scientifiques s'adressant principalement aux enseignants chercheurs linguistes. Nous souhaitons récolter un corpus suffisamment large pour qu'il soit représentatif de la langue, et afin qu'il permette sa sauvegarde sous forme de ressource linguistique. Cette recherche fait intervenir deux institutions : le département de langue et culture de Tizi-Ouzou et la section linguistique de la Faculté de lettres de l'université de Lausanne. Aucun moyen financier spécifique n'accompagne ce projet mais ce dernier a été intégré dans le laboratoire de recherche « Aménagement et enseignement de la langue amazighe » agrée en 2009.

Les objectifs :

Le premier objectif est de mettre à disposition de linguistes une ressource linguistique ce qui implique des conséquences sur la manière de définir les métadonnées et les annotations. Cette recherche est aussi une occasion de documenter le kabyle parlé dans toutes ses variétés, sous toutes ses formes géographiques. Son intérêt réside aussi du fait que cette ressource linguistique sera accessible via le web. Ainsi, on peut ajouter que cette banque de corpus n'a pas pour objectif le TAL ou le TIC mais un outil aussi complet possible (métadonnées, annotations, étiquetage…) pour des linguistes qui pourraient s'intéresser à un ou des élément(s) de recherche.

La création d'un corpus oral, se fait sur la base l'article de Jacobson (2002), chercheur au LACITO (Laboratoire de langues et civilisations à tradition orales). Nous intégrerons l'écrit en utilisant la notation usuelle du kabyle. Les corpus constitués, nous les écrivons en notation usuelle et les retranscrivons en transcription phonétique (API) (Annexe transcription). Cette opération étant faite, nous y ajoutons des métadonnées qui permettront d'identifier nos données et les

décrire (date, langue etc.). Nous nous basons sur les recommandations d'OLAC pour le codage des métadonnées (LACITO,

http://lacito.vjf.cnrs.fr/archivage/index.htm) même si d'autres modèles (ALAVAL,http://www2.unine.ch/dialectologie/page9353.html, CRDO, http://crdo.risc.cnrs.fr/exist/crdo/ et http://crdo.up.univ-aix.fr/) sont aussi intéressants.

La conservation des données se fera grâce à des copies et à la numérisation (transformation en ressource linguistique informatisée). En effet, comme le rappelle Jacobson (2002), le mode de représentation digital a l'avantage d'être répandu, facile d'emploi et a la capacité de mieux conserver les données. Nous utiliserons un codage sans compression pour nos données audio, ce qui semble plus adapté pour l'archivage à long terme.

Le travail sur le terrain :

Pour atteindre notre but nous enregistrons des corpus de locuteurs monolingues. Ces corpus sont recueillis par nos étudiants de licence de notre département. Ceci a un double objectif : - cibler toutes les régions de la Kabylie grâce à eux qui proviennent des quatre coins de notre terrain d'enquête. – compléter la formation de nos étudiants. Des consignes strictes sont données aux enquêteurs : Faire transcrire le même corpus par deux étudiants, indépendamment l'un de l'autre. Un membre de l'équipe comparera ensuite ces deux transcriptions pour repérer d'éventuelles écarts récurrents (par exemple variation fréquente entre [k] et [t], entre occlusive et spirante etc.) qui peuvent être l'indice de difficultés. Contrôler toutes les transcriptions faites par les étudiants indépendamment par deux
membres de l'équipe (avec réécoute de l'enregistrement simultanément) et la faire évaluer grossièrement (par exemple: Très bon - Bon - Suffisant - Insuffisant). On comparera ensuite les évaluations données et on réexaminera les transcriptions pour lesquelles les évaluations diffèrent de façon importante (de plus d'un degré). On réexaminera également toutes les transcriptions jugées insuffisantes par un évaluateur au moins pour décider de celles qui devraient être écartées comme trop fautives et refaites. On identifiera clairement quels étudiants ont transcrit quels corpus, quels membres de l'équipe l'ont contrôlé et conserver cette information (ce seront des métadonnées importantes). Il pourrait être utile d'avoir des informations de type sociolinguistique sur les étudiants qui transcrivent...

Nous avons établi pour chaque locuteur une fiche de collecte (Annexe 1) où doivent apparaître les métadonnées préalablement définies. Pour compléter ces données, nous avons établi des listes de mots (Annexe 2) en fonction de plusieurs paramètres dont les différents champs sémantiques que nous soumettons dans les divers points d'enquête.

Choix technologiques :

Nous avons opté pour l'adoption de standarts (OLAC ; xml) et des logiciels autant que possible gratuits, open-source et multi-plateformes (Windows-Mac_OSX-Linux).

Le traitement et l'informatisation des corpus oraux supposent un certain nombre d'outils théoriques et de techniques qu'on devait maîtriser. Le premier point est la définition des métadonnées. La question des métadonnées commence à se poser sérieusement lorsque se multiplient les ressources linguistiques informatisées et potentiellement accessibles en ligne. Il s'agit de se mettre d'accord sur des descripteurs qui permettront ensuite une recherche efficace dans un catalogue qui renverra aux ressources elles-mêmes.

Dans la constitution d'un système de métadonnées pour des données - ou "ressources" - linguistiques (enregistrements audio ou video, photos, transcriptions, annotations), différents niveaux peuvent être considérés:

- *Description générale de la ressource linguistique (langue, variété, date de recueil, genre…)*
- *Description des traits spécifiques de la ressource linguistique (date, lieu, enquêteur, informateur, moyens techniques, fichiers (noms, types, localisation...)*

Pour notre projet cela nous concerne

1) puisque l'un des objectifs, dans l'avenir, est de publier les informations sur les ressources construites pour permettre à d'autres chercheurs de savoir qu'elles existent et, le cas échéant, d'y accéder. (Mais rendre publiques les métadonnées n'impliquent pas obligatoirement de rendre l'accès à ces données également libre)

2) Comme il est prévu un grand nombre de corpus élémentaires (=enregistrements ou sessions...) il faut alors, de toutes façons, se construire un système de métadonnées pour retrouver rapidement un sous-ensemble de données. Alors autant le construire de façon à ce qu'il soit compatible avec un système standardisé.

En relation avec ces structures de métadonnées des logiciels capables de les utiliser ont été développés (OLAC, IMDI…)

Pour notre part nous avons choisi d'utiliser OLAC (http://linguistlist.org/olac/index.html)

L'OLAC a élaboré son système de métadonnées pour la description de ressources linguistiques. Il est simple et assez général, mais la formalisation d'un mécanisme

d'extensions permet d'être plus spécifique.

Pour notre recherche l'examen, même rapide, de ces systèmes de métadonnées a eu le mérite de nous permettre de contrôler que rien d'essentiel n'a échappé à notre projet de "fiche de collecte". On voit ainsi, par exemple, que cette fiche ne permet pas de décrire le genre de données recueillies: soliloque, conversation, réponses orales à des questions, poèmes etc...

D'autre part les notations de lieux (d'enquête, de naissance etc) devraient être précisées par une indication longitude/latitude en raison du grand nombre de noms de lieux identiques - donc ambigus - en Kabylie.

Actuellement, nous sommes arrivés à 700 points d'enquête, et 400 enregistrements de 20mn chacun pour la plupart transcrits (Annexe 3 : exemple de corpus). Nous avons établi une « carte exemple » d'un certain nombre de points d'enquête (Annexe 4).

Nous avons, pour le moment utilisé Google Earth pour la représentation spatiale de ces points d'enquête ; La définition des coordonnées de ces points (longitude et latitude) n'a pas été une tâche facile. En effet, les toponymes présentent une grande variation dans le temps et dans l'espace. Il nous arrive de ne pas pouvoir situer exactement un point d'enquête sur la carte parce le nom a changé ou a été transformé. En effet, les diverses sources (cartes topographiques, enquêtes de Basset, documents administratifs fournis par la Wilaya) présentent parfois, des variations importantes dans les toponymes et ceci est une difficulté supplémentaire à surmonter quand on passe à une représentation cartographique.

Enrichissement des données :

La première opération indispensable pour passer de corpus oraux au corpus écrits est la préparation d'un clavier qui pourrait nous faciliter l'utilisation des caractères spécifiques du kabyle. Pour ce faire nous sommes partis des conventions d'écriture de l'INALCO (http://www.inalco.fr/crb/pages_htmel/tableau_prononciation_kab.html) et UNICODE pour élaborer ce clavier.

Unicode c'est fantastique parce qu'on peut utiliser des dizaines de milliers de caractères dans une seule police...

Mais Unicode c'est infernal parce qu'on peut réaliser la même lettre de plusieurs façons différentes et que ces différences, si elles ne sont pas toujours facilement perçues par l'oeil humain, sont un abîme pour un ordinateur.

Le problème se pose pour les caractères complexes (notation des emphatiques par exemple) qui peuvent exister comme caractères uniques en quelque sorte pré-

construits et occupant une position dans la grille Unicode ou bien être produit par l'association de deux caractères: une lettre et un signe diacritique.

Or les programmes informatiques vont traiter différemment ces deux situations. Les logiciels permettant d'établir des listes de fréquence ou des concordances fonctionnent correctement lorsque les caractères complexes sont codés par des caractères uniques mais ne savent pas traiter le cas où ils sont formés par l'association de deux caractères.

La règle à appliquer est donc la suivante: si c'est possible, écrire un caractère complexe en utilisant un caractère unique et non pas en combinant un caractère littéral et un caractère diacritique.

C'est pourtant exactement l'inverse que propose le site edition.berbere...

(http://edition.berbere.free.fr/tables_saisie_berbere_utf-8_01.html).

Les propositions qui figurent dans le tableau suivant respectent la règle ci-dessus

Pour les consonnes labiovélaires (pas reprises ici) il n'y a pas de caractères uniques dans Unicode. La proposition de l'INALCO - postposition de ° - reste donc la plus simple puisque ° en exposant se trouve directement sur tous les claviers.

	Bloc	Code		Bloc	Code
a	Latin de base	0061	A	Latin de base	0041
b	Latin de base	0062	B	Latin de base	0042
c	Latin de base	0063	C	Latin de base	0043
č	**Latin étendu-A**	**010D**	**Č**	**Latin étendu-A**	**010C**
d	Latin de base	0064	D	Latin de base	0044
ḍ	**Latin étendu suppl.**	**1E0D**	**Ḍ**	**Latin étendu suppl.**	**1E0C**
e	Latin de base	0065	E	Latin de base	0045
f	Latin de base	0066	F	Latin de base	0046
g	Latin de base	0067	G	Latin de base	0047
ǧ	**Latin étendu-B**	**01E7**	**Ǧ**	**Latin étendu-B**	**01E6**

(NPC avec ğ et Ğ [diacritique : brève !] de Latin étendu-A 011F et 011E)

h	Latin de base	0068	H	Latin de base	0048
ḥ	**Latin étendu suppl.**	**1E25**	**Ḥ**	**Latin étendu suppl.**	**1E24**
i	Latin de base	0069	I	Latin de base	0049
j	Latin de base	006A	J	Latin de base	004A
k	Latin de base	006B	K	Latin de base	004B
l	Latin de base	006C	L	Latin de base	004C
m	Latin de base	006D	M	Latin de base	004D
n	Latin de base	006E	N	Latin de base	004^{E}
ɣ	**Extensions IPA**	**0263**	**Ɣ**	**Latin étendu-B**	**0194**

(Attention ! autre possibilité : bloc Grec et Copte avec le couple γ 03B3 *pour la minuscule et* Γ 0393 *pour la capitale. On pourrait réserver ces caractères, si nécessaire, à la notation d'une réalisation spirante d'un /g/)*

q	Latin de base	0071	Q	Latin de base	0051
r	Latin de base	0072	R	Latin de base	0052
ṛ	**Latin étendu suppl.**	**1E5B**	**Ṛ**	**Latin étendu suppl.**	**1E5A**
s	Latin de base	0073	S	Latin de base	0053
ṣ	**Latin étendu suppl.**	**1E63**	**Ṣ**	**Latin étendu suppl.**	**1E62**
t	Latin de base	0074	T	Latin de base	0054
ṭ	**Latin étendu suppl.**	**1E6D**	**Ṭ**	**Latin étendu suppl.**	**1E6C**
ţ	**Latin étendu-A**	**0163**	**Ţ**	**Latin étendu-A**	**0162**
u	Latin de base	0075	U	Latin de base	0055

w	Latin de base	0077	W	Latin de base	0057
x	Latin de base	0078	X	Latin de base	0058
y	Latin de base	0079	Y	Latin de base	0059
z	Latin de base	007A	Z	Latin de base	005A
ẓ	**Latin étendu suppl.**	**1E93**	**Ẓ**	**Latin étendu suppl.**	**1E92**
ɛ	**Extensions IPA**	**025B**	**Ɛ**	**Latin étendu-B**	**0190**

(Attention ! d'autres possibilités seraient envisageables, p.ex. bloc Grec et Copte...)

Généralisation à la notation de la spirantisation.

Le principe consistant à préférer systématiquement l'utilisation d'un caractère unique sur l'association de deux caractères est également préférable pour les autres niveaux de transcription. Ainsi, pour la notation des spirantes, dans une transcription phonétique large, si l'on décide d'adopter la convention du trait souscrit (suscrit sur g ou G) plutôt que le recours aux caractères de l'API, il vaudra mieux utiliser les caractères qui apparaissent dans le bloc Latin étendu supplémentaire plutôt que de combiner un caractère avec le diacritique « trait souscrit » (Unicode 0320).

Pour écrire le kabyle, en plus des lettres habituelles on a besoin:

- des lettres: ɣ et ɛ
- des lettres diacritées: c et g avec caron (appelé encore: hacek, chevron, antiflexe, accent hirondelle, v suscrit), d, h, r, s, t, z avec point souscrit, t cédille.

Tous ces caractères doivent être disponibles en lettres minuscules et en lettres capitales (majuscules).

Toutes ces lettres sont prévues, précomposées, dans divers blocs Unicode. Les codes correspondants sont indiqués ci-dessus.

Pour permettre la saisie de ces lettres sans exiger de trop gros efforts de mémorisation et éviter des conflits avec des combinaisons de touches prédéfinies par le système ou par d'autres programmes (Word par exemple), la solution

générale retenue consiste à définir une "touche morte", au fonctionnement analogue à la touche de l'accent circonflexe ou du tréma. On presse la touche morte puis la touche correspondant portant un caractère simple (dit "de base") et on obtient le caractère spécial voulu.

La touche retenue comme touche morte est celle qui, sur le clavier suisse romand, porte les signes < et >.

Le ɣ et le ɛ s'obtiennent avec la touche morte suivie des touches y et e (les caractères de base les plus proches par leur forme).

Les ḍ , ḥ , ṛ , ṣ , ṭ et ẓ avec la touche morte suivie des caractères de base correspondants d, h, r, s, t et z.

Pour ṭ la touche morte est suivie de la touche x ("i*ks")*.

Les lettres capitales s'obtiennent normalement en combinant la touche morte avec la touche shift (majuscule).

Les caractères < et > restent disponibles: il suffit de les taper après la touche morte: deux pressions successives sur la touche < donne < ou > si la touche shift est pressée.

Concrètement, et suivant Sur Macintosh (avec clavier Français-Suisse), ou sur PC, il faut suivre les opérations suivantes :

Sur Macintosh (avec clavier Français-Suisse)

- installer le fichier +kabyle.keylayout (créé avec le logiciel gratuit Ukelele cf. scripts.sil.org/ukelele) dans le dossier Keyboard Layouts qui se trouve dans le dossier Bibliothèque (ou: Library) de l'utilisateur (ou de l'ordinateur iMac). (Si le dossier Keyboard Layouts n'existe pas il faut le créer dans le dossier bibliothèque, en lui donnant exactement ce nom);

- redémarrer l'ordinateur;

- ouvrir les Préférences Système... (menu Pomme) et ensuite International; cliquer sur l'onglet Menu Saisie, rechercher le clavier +kabyle et cocher la case à gauche (Activé);

- dans la barre en haut de la fenêtre, à droite, cliquer sur le drapeau qui symbolise le clavier (combinaison des drapeaux suisse et français) et sélectionner le clavier +kabyle qui doit se trouver en dessous.

Le clavier +kabyle est désormais accessible et toute application utilisant une police

Unicode assez complète (comme Doulos SIL) permettra d'obtenir les caractères spécifiques nécessaires avec la touche morte.

Mais il y a un logiciel gratuit, Microsoft Keboard Layout Creator, accessible ici: http://www.microsoft.com/globaldev/tools/msklc.mspx, qui permet de reconfigurer un clavier et, notamment, de créer une touche morte.

Toutefois l'utilisation de ce programme exige l'installation préalable de l'environnement de programmation .NET (.NET Framework) à télécharger ici :

http://www.microsoft.com/net/Download.aspx

Un mode d'emploi en français ci-joint (MKLC_fr.pdf; extrait de http://llacan.vjf.cnrs.fr/fichiers/manuels/Internet/SaisieClavier.pdf) permet de se débrouiller assez facilement. Il faut simplement corriger ce qui est dit sur l'installation du clavier dans Windows:

- le fichier .msi est le fichier composé du nom du clavier et de l'abréviation de la famille du processeur (le plus souvent i386). Mais il y a un fichier de Setup qui doit se charger d'installer la bonne version. Attendre le message: Installation complète. Ce n'est pas immédiat.

- c'est le panneau de configuration Options régionales (et non Clavier) qui, sous Windows XP en tout cas, permet d'installer et d'activer le nouveau clavier.

Bien entendu il est possible de choisir n'importe quelle touche comme touche morte, pas seulement le <.

Extension envisageable:

Si on le souhaite, on peut ajouter d'autres caractères, comme par exemple le ə (<+a). La difficulté consiste à attribuer les caractères supplémentaires à une touche présentant, si possible, un certain rapport, pour éviter un effort de mémoire. Mais on pourrait parfaitement, par exemple, définir une autre touche morte pour entrer des caractères de l'alphabet phonétique utilisés dans une transcription phonético-phonologique. On pourrait avoir, par exemple, avec $ comme touche morte $+t donnant θ, $+d donnant δ etc.

Sites cités:

- pour télécharger Ukelele, logiciel de configuration du clavier pour Macintosh:

http://scripts.sil.org/ukelele

- pour télécharger Microsoft Keboard Layout Creator, logiciel de configuration du

clavier pour PC:

http://www.microsoft.com/globaldev/tools/msklc.mspx

- pour télécharger l'environnement de programmation .NET pour Windows, s'il n'est pas installé: http://www.microsoft.com/net/Download.aspx

Enrichissement des données :

L'enrichissement des données par un certain nombre de logiciels présuppose la mise en place d'un certain nombre de concepts qui pourraient nous aider dans la segmentation des corpus en unités (énoncés, phrase…) et l'étiquetage linguistique (morphosyntaxique). L'un des points qui nous intéressent est la relation phrase/prosodie/segmentation.

Philippe Martin (1981, 2002, 2010) définit assez clairement les concepts qui nous intéressent pour notre problématique. Ainsi, pour lui, le mot prosodique est l'unité prosodique minimale contenant un seul mot accentué. Cela correspond, généralement, au syntagme. Ceci explique la composition du groupe prosodique de mots prosodiques. Quant à la phrase prosodique, toujours d'après Philippe Martin, elle indique la courbe mélodique phrastique, dépendant de la modalité de la phrase (déclarative, interrogative etc.).

Dans cette perspective, la phrase prosodique n'est qu'une suite de mots délimitée par deux pauses importantes (initiale et finale) et caractérisée par une intonation qui varie avec le type de phrase (assertive, interrogative, injonctive).

Pour la définition de la phrase et de l'énoncé, le Dictionnaire de la linguistique de Georges Mounin (2004 : 262) stipule "Beaucoup d'usages linguistiques tiennent *énoncé* et *phrase* pour des termes synonymes. Mais on a intérêt à opposer les phrases (unités de langue) aux énoncés (unités ou exemples de parole), l'énoncé étant ce qui est donné dans le matériau non analysé".

Un énoncé est "tout segment de la chaîne parlée, compris entre deux interruptions nées soit du silence, soit du changement de locuteur, et qui n'a pas encore été identifié ou analysé en phrases" (G. Mounin, 2004 : 125).

La phrase est définie par A.Martin (1991 : 131) comme une séquence *«dont tous les éléments se rattachent à un prédicat unique ou à plusieurs prédicats coordonnés »*. Pour la syntaxe, il déclare (1985 :13) «S'il est un point sur lequel peuvent tomber d'accord les linguistes contemporains, à quelque école qu'ils se rattachent, c'est qu'appartient à la syntaxe l'examen de la façon dont les unités linguistiques douées de sens se combinent, dans la chaîne parlée, pour former des énoncés (…) c'est-à-dire la façon d'ordonner des mots pour former des phrases». Et l'objet de la syntaxe est «d'exprimer par quels moyens les rapports qui existent entre les éléments d'une expérience(…) peuvent être marqués dans une succession

d'unités linguistiques de manière que le récepteur du message puisse reconstruire cette expérience» (Ibid., 2-8, p.16)

Toujours pour notre étiquetage linguistique, nous avons un élément aussi important que la phrase et l'énoncé, à savoir le syntagme pour qui la définition de A.Martinet est tout à fait indiquée puisqu'il le définit (Martinet, 1980, 4-13, p.112) comme « toute combinaison de monèmes dont les rapports mutuels sont plus étroits que ceux qu'ils entretiennent avec les autres éléments de l'énoncé, plus, éventuellement, le monème fonctionnel qui rattache cette combinaison au reste de l'énoncé»

Dans un énoncé complexe, on trouve donc un énoncé minimum qui se compose généralement d'un prédicat et d'un sujet (expansion obligatoire) et les expansions. Le prédicat est l'élément irréductible de l'énoncé.

Donc, une phrase est un énoncé dont tous les éléments se rattachent à un prédicat unique ou à plusieurs prédicats coordonnés en tenant compte des pauses importantes (initiale et finale) de l'intonation qui varie avec le type de phrase (assertive, interrogative, injonctive).

Deux logiciels gratuits et libres d'accès sont utilisés : PRAAT(http://www.fon.hum.uva.nl/praat/download_win.html)

et JAXE (http://sourceforge.net/projects/jaxe/)

JAXE :

Un langage XML est défini de façon formelle, de manière à permettre la vérification automatique de la syntaxe. Cette description formelle comprend les noms des éléments du langage, les imbrications possibles entre les éléments, l'ordre autorisé des éléments, et leurs attributs (les attributs étant optionnels ou obligatoires).

Jaxe facilite la création des documents XML en utilisant les règles du langage pour proposer des éléments à insérer là où c'est possible. Cela permet de créer des documents *valides* (c'est à dire se conformant aux règles du langage) beaucoup plus facilement qu'avec un simple éditeur de texte.

En plus des fichiers décrivant les langages XML (les `schémas`), Jaxe utilise des *fichiers de configuration* qui définissent la barre de menus et la façon d'afficher les éléments du langage. Ces fichiers se trouvent dans le répertoire `config`, et leur nom se termine en `_config.xml`.

La composante « Métadonnées »

Chaque enregistrement est accompagné d'une « fiche de collecte » qui le décrit. Cette fiche de collecte :

- Sera intégralement transcrite sous la forme d'un fichier xml (éléments : Divers, Enquêté, Collecteur, Debriefing, Autres infos ; sous-éléments : les différentes lignes de la fiche), auquel il sera référence dans le document élémentaire ;

- Constituera la source des métadonnées incluses dans le document élémentaire.

  Le choix des métadonnées retenues pour accompagner directement chaque document élémentaire se fonde sur les standarts reçus (Dublin Core (DC) et OpenLangage Archives Community (OLAC) et suit d'assez près les recommandations du Centre de Ressources pour la Description de l'Oral (CRDO, CNRS). On renonce cependant à noter les caractéristiques constantes de nos documents : la langue étudiée (le kabyle) et la langue d'étude (le français)

- On donne, sous-l'élément Métadonnées, la liste des sous-éléments (= représentation de la structure hiérarchique)

- On définit ensuite, comme des éléments distincts, en dehors de la spécification de l'élément Métadonnées, chacun de ces sous-éléments (= représentation des composantes de la structure). Ces sous-éléments de l'élément Métadonnées sont donc des éléments et peuvent à leur tour se composer de sous-éléments.

Cette représentation est donnée en Annexe 5

La composante de l'élément Données

La première composante est la Phrase qui va être analysée en une succession de parties du discours et qui peut-être glosée (traduction juxtallinéaire), transcrite, en phonétique ou en phonologie, traduite. Elle est également liée à un élément sonore.
Les éléments de glose, de transcriptions, de traduction et de lien avec le signal audio caractérisent également les différentes parties du discours. C'est pourquoi on les réunit en un « ensemble » (nommé, dans cet exemple, « formes »).
Les parties du discours, ainsi que les attributs qui les caractérisent, sont déterminées par les linguistes berbérisants du groupe de recherche.
Pour que le fichier de description de la structure soit accepté par Jaxe, il faut encore indiquer un élément racine de l'arborescence hiérarchique. Dans notre exemple ce sera l'élément Document_kabyle.
L'illustration est en Annexe 6, 7, 8, 9.

Application avec PRAAT :

PRAAT () est exploité en analyse acoustique. En créant de nombreuses tires, on arrive aligner le signal temporel, le sonagramme, la notation usuelle, le découpage en unités préalablement définies ou étiquetage linguistique (racines, schèmes, syntagmes…) (Annexe 10). Des scripts sont également utilisés à des fins de segmentation en énoncés par exemple. Evidemment toute la problématique de la

définition de l'énoncé en ce qui concerne l'oral est difficilement maîtrisable. Pour notre part, les pauses sont prises comme indicateur de séparations d'énoncés (Annexe 11, 12). Evidemment PRAAT a aussi la qualité d'aligner son/transcription.

Annexe 1 :

Fiche de collecte

1. divers		
date de collecte :	2009	
lieu :	Tigzirt (Tasalast et Tamda Ouguemoune)	
support de l'enregistrement :		
durée de l'enregistrement :	Environ 45 minutes	
lieu de l'enregistrement :	Au bord de la mer	
sujet de l'enregistrement :	Poissons, animaux de la mer,	
Y avait-il un public ?	Non	
Référence		
2. enquêté		
(Nom :)		
Date de naissance :	L'un est né en 1934 / l'autre est né en 1977	
Sexe :	Hommes	
Village d'origine :	Tigzirt	
Tribu :	Iflissen	
Domicile actuel (village, région):	Tigzirt	
Dialecte parlé, (nom donné par le locuteur à son parler)	Kabyle	
Autre (s) langue (s) parlée (s) :	Kabyle, arabe	
(Au travail :)	?	

(À la maison :)	Kabyle	
Séjour (s) à l'étranger	Non	
Durée du/des séjour(s)	?	
Scolarité et formation	Niveau CEM celui qui est né en 1977	
Langue(s) de l'enseignement reçu :		
Profession :	Chasseurs marins	
Personne(s) ayant joué un rôle dans l'apprentissage linguistique (par exemple son père, sa mère, personne avec qui le locuteur a passé son enfance)		
- lien de parenté, relation avec la personne :	Non	
- lieu d'origine :		
- scolarité (et langues d'enseignement) :		
situation familiale (mariage(s), enfants) :	Marié (celui qui est né en 1934), célibataire (celui qui est né en 1977)	
langue (s) parlée (s) par le conjoint :	Kabyle	
attitude du locuteur par rapport à sa langue et à sa façon de parler :	Fière vis-à-vis du kabyle, leur parler différent au reste de la Kabylie ;	
3. Collecteur		
nom, prénom:	Oumaouche Omar	

langue (s) parlée (s) :	Kabyle, arabe, français	
origine :	Tigzirt	
relation enquêteur-enquêté :	?	
4. Debriefing		
conscience du micro :		
attitude du locuteur par rapport à l'enregistrement :		
attitude du locuteur par rapport à l'entretien, aux questions posées…		
5. Autres infos		

Annexe 2 :

mot (en français)	Parler 01 : Aglala / Beni Zmenzer	Parler 02 : Isseradjène / Boudjima
Champignon	*Tireγlin Racine : RΓ L Schème: $tic_1ec_2c_3ic_4$	*Tireγla Racine : RΓ L Schème : $tic_1ec_2c_3a$
Petits pois	*Tajijbant R : JBN S : $tac_1ic_2c_3ac_4t$	*Tajilbant R : JLBN S : $tac_1ic_2c_3ac_4t$
Pin Silvestre	*Tazumbit R : ZMB S : $tac_1uc_2c_3it$	*Tazumbilt R : ZMBL S : $tac_1uc_2c_3ic_4t$
Citronnier	*Tilimet R : LM S : tic_1ic_2et	*Talimet R : LM S : tac_1ic_2et
Lentisque	*Imidek R : (m) DK S : $ic_1ic_2ec_3$	*Tidekt R : DK S : ti c_1ec_2t
Lentilles	*Laɛ des R: □DS S : $c_1ac_2c_3ec_4$	*Lɛ eḍ s R : □Ḍ S S : $c_1c_2ec_3c_4$
Eucalyptus	*Karitus R : KRTS S : $c_1ac_2ic_3u\ c_4$	*Akalatus R : KLTS S : $ac_1ac_2ac_3uc_4$
Chêne liège	*Akerruc R : KRC S :a $c_1eC_2uc_3$	*Ikirruc R : KRC S : $ic_1iC_2uc_3$
Gland	***A**ḥ **ellu**ḍ ***R :*** Ḥ ***L***Ḍ ***S : $ac_1eC_2uc_3$***	****Abellu**ḍ **R : BL**Ḍ **S : $ac_1eC_2uc_3$***
Rue	*Awermi R : WRM S : $ac_1ec_2c_3i$	*Lfenğla R : LFĞL S : $c_1c_2ec_3c_4c_5a$

Rosier	*Tiɛ fart : R : Γ FR S : $tic_1c_2ac_3t$	*Taɛ fart R : Γ FR S : $tac_1c_2ac_3t$
Palmier	*Tazdayt R : ZDY S : $tac_1c_2a\ c_3t$	*Tazanet R : ZN S : tac_1ac_2et
Citrouille	*Taxsayt R : XSY S : $tac_1c_2a\ c_3t$	*Taxsakt R : XSK S : $tac_1c_2ac_3t$
Luzerne	*Ikeffis R : KFS S : $ic_1eC_2ic_3$	*Ikeffil R : KFL S : $ic_1eC_2ic_3$
Figuier de barbarie	*Lkermus R :(L) KRMS S :$c_1c_2e\ c_3c_4uc_5$	*Lkermus R: (L) KRMS S :$c_1c_2ec_3c_4uc_5$
Mûres sauvages	* Tinijwal R : NJWL S : $tic_1ic_2c_3ac_4$	*Tizwal R: ZWL S : $tic_1c_2ac_3$
	* Timendekrar R : MNDKR S : $tic_1ec_2c_3ec_4c_5ac_6$	*Tiferkekkay R : FRKY S : $tic_1ec_2c_3eC_4ac_5$
	*Timeccucin R : MC S :$tic_1eC_2uc_3ic_4$	*Aɛersiwen R : Γ RṢ W S : $ac_1ec_2c_3ic_4ec_5$
Lait	*Ayefki R : YFK S : $ac_1ec_2c_3i$	*Ifki R : FK S : i c_1c_2i
Frère	*Xuya R : XY S : c_1uc_2a	*Ugma R : GM S : uc_1c_2a
L'argent donné pour la fiancée	*Tiẓri R: ẒR S : ti c_1c_2i	*Tiẓri R : ẒR S : tic_1c_2i
Filles	*Tullas R : LS S : tuC_1ac_2	*Tiḥdayin R :ḤDY S :$tic_1c_2ac_3ic_4$
Amas de bois qui sert à cuire la poterie	*Uγud R : ΓD	*Uγud R : ΓD

	S : u c_1uc_2	S : uc_1uc_2
Cruche	*Acmux R : CMX S : $ac_1c_2uc_3$	*Asagem R : GM S : $ac_1ac_2ec_3$
Chapelet de morceau de viande	*Iceddiwen R: CDW S : i $c_1C_2ic_3ec_4$	*lmeck R : (L) MCK S : $c_1c_2ec_3c_4$
Un spécialiste pour circoncire les garçons	*aḥeğğ***am** ***R :*** ḤĞ***M*** ***S : $ac_1eC_2ac_3$***	*lemɛellem R : (L) ƐLM S : $c_1ec_2c_3eC_4ec_5$
Les œufs à la semoule	*Timcewwect R : MCW S : $tic_1c_2eC_3ec_4t$	*Tabeγrirt R : BΓR S : $tac_1ec_2c_3ic_4t$
Bébé	*Llufan R : LFN S : $C_1uc_2ac_3$	*Agrud R : GRD S : $ac_1c_2uc_3$
Crêpes	*Lemsemmen R : (L) MSMS S :$c_1ec_2c_3eC_4ec_5$	*Aḥ***eddur** ***R :***Ḥ***DR*** ***S : a $c_1eC_2uc_3$***
Petite fille	*Taqcict R : QC S :$tac_1c_2ic_3t$	*Tagruḍt R : GRḌ S : $tac_1c_2uc_3$
Faire purifier, circoncire clarifier	*Sḍehren R :ḌHR S :$c_1c_2ec_3c_4ec_5$	*Zeyynen R : ZYN S :$c_1e\,C_2c_3ec_4$
Faire manger	*Aceĕĉi R : C S : ac_1eC_2i	*Aseĕĉi R : C S : ac_1eC_2i
Enfant	*Aqcic R : QC S : $ac_1c_2i\,c_3$	*Aqcic R : QC S : $ac_1c_2ic_3$
A ce moment là	*Imir-n R : MR S : $ic_1ic_2c_3$	*Imir R : MR S : ic_1ic_2
Souhait	*Saɛd R : SƐd	*Henni R : HN

	S : c_1a c_2c_3	S : c_1eC_2i
Je lui dois, elle leur doit	*Tettalas R : LS S : te$C_1ac_2ac_3$	*Tettaras R : RS S : te$C_1ac_2ac_3$
Peuvent	*Zemren R ZMR S : $c_1ec_2c_3ec_4$	*Waɛan R : WƐ S : $c_1ac_2ac_3$
Haut de robe qui bouffe en poche au dessus de la ceinture	*Icimmi R : CM S : $ic_1i\,C_2i$	*Iciwi R : CW S : ic_1ic_2i
La cuise	*Taγma R : ΓM S : tac_1c_2a	*Taqesbuḍt R: QSBḌ S : $tac_1ec_2c_3uc_4$
Celles qui roule la semoule avec les mains dans un grand plat pour la préparation du couscous	*Tifettalin R : FTL S : $tic_1eC_2a\,c_3ic_4$	*Tineffalin R : NFL S : $tic_1e\,C_2a\,c_3ic_4$
Etre d'accord	*Mseqbalen R :(MS) QBL S : $c_1c_2ec_3c_4ac_5ec_6$	***mru**ḍ**an*** R: (M)RḌ S :$c_1c_2uc_3ac_4$
Marie	*Isli R : SL S : ic_1c_2i	*Isli R : SL S : ic_1c_2i
Ce qu'il faut	*Ilaqen R : LQ S : $ic_1ac_2ec_3$	*Ilezmen R : LZM S : $ic_1ec_2c_3ec_4$
Faire les youyous	*Siγret R : ΓRT S : $c_1ic_2c_3ec_4$	*Seγret R : ΓRT S : $c_1ec_2c_3ec_4$
Beignets	*Lesfenğ R : (L) SFNĞ S : $c_1ec_2c_3ec_4c_5$	*Lexfaf R (L) XF S : $c_1ec_2c_3ac_4$
Insectes	*Ibaɛɛac R : BƐC S : $ic_1aC_2ac_3$	*Ibelɛac R : BLƐ S : $ic_1ec_2c_3ac_4$
	*Aqrur	*Agrud

Enfant	R : QR S :$ac_1c_2uc_3$	R : GRD S : $ac_1c_2uc_3$
Les enfants	*Arrac R : RC S : aC_1ac_2	*Igerdan R : GRD S : $ic_1ec_2c_3ac_4$
Ensemble	*Ilkelli R : LKL S : $ic_1c_2eC_3i$	*Urkelli R : RKL S :$uc_1c_2eC_3i$
Faire partie du cortège qui chercher la marie	*Iɛerrasen R :ƐRS S : $ic_1eC_2ac_3ec_4$	*Iqeffafen R : QF S : $ic_1eC_2ac_3ec_4$
Toute petite	*Taṭuṭaḥt R :ṬḤ S : $tac_1uc_2ac_3t$	*Tamecṭuḥt R :MCṬḤ S : $tac_1ec_2c_3uc_4t$
Bébé	*Ṣṣebyan R :ṢBY S : $C_1ec_2c_3ac_4$	*Agrud R : GRD S : $ac_1c_2uc_3$
Veau	*Aɛejmi R : ƐJM S : $ac_1ec_2c_3i$	*Agenduz R : GNDZ S : $ac_1ec_2c_3uc_4$
Achille gold	*Γlilu R :ΓL S :$c_1c_2ic_3u$	*Qlilu R : QL S : $c_1c_2ic_3u$

Cigale	*Zḍeğ R : ZḌĞ S :$c_1c_2ec_3$	*Tejḍeč R : JḌČ S : $tec_1c_2ec_3$
Chouette	*Timiɛruft R : MƐRF S : $tic_1i\,c_2c_3uc_4t$	*Imiɛruf R : MƐRF S : $ic_1i\,c_2c_3uc_4$
Hirondelle	*Tifilellest R : FLS S : $tic_1ic_2eC_3ec_4t$	*Tifirellest R : FRLS S : $tic_1ic_2eC_3ec_4t$
Papillon du jour	*Timecriwect R : MCRWC S : $tic_1ec_2c_3ic_4ec_5t$	*Aferṭiṭu R : FRṬ S : $ac_1ec_2c_3ic_4u$

Papillon de nuit	*Aferṭeṭu R : FRṬ S : $ac_1ec_2c_3ec_4u$	*Aferṭiṭu R : FRṬ S : $ac_1ec_2c_3ic_4u$
Singe	*Iddew R : DW S : iC_1ec_2	*Ibekki R : BK S : ic_1eC_2i
Chauve-souris	*Ṭirellil R : ṬRL S : $c_1ic_2eC_3ic_4$	*Itirelli R : TRL S : $ic_1ic_2eC_3i$
Scorpion	*Tiγirdemt R : ΓRDM S : $tic_1ic_2c_3ec_4t$	*Tawejjiḍt R : WJḌ ***S : $tac_1eC_2ic_3$***
Tordeuse	*Aburebbu R : RB S :$abuc_1C_2u$	*burebbu R :RB S :buc_1C_2u

Taon	*Taggent R : GN S : taC_1ec_2t	*Aggen R : GN S : aC_1ec_2
Faucon	*Afalku R : FLK S : $ac_1ac_2c_3u$	*Lbaz R : LBZ S : $c_1c_2ac_3$
Bon plat	*Taḥluqt R : ḤLQ S : $tac_1c_2uc_3t$	*Tarzeft R : RZF S : $tac_1c_2ec_3t$
Jeune pousse	*Issegmi R : SGM S : $iC_1ec_2c_3i$	*Aguḍem R : GḌM S : $ac_1uc_2ec_3$
Renard	*Izirdi R : ZRD S : $ic_1ic_2c_3i$	*Akɛab R : KƐB S : $ac_1c_2ac_3$
Vache	*Tafunast R : FNS S : $tac_1uc_2ac_3t$	*Tuwmat R : WM S : tuc_1c_2at
	*Taselluft	*Taselluft

Puce	R : SLF S : $tac_1eC_2uc_3t$	R : SLF S : $tac_1eC_2uc_3t$
Chevale	*Aɛewdiw R : ƐWDW S : $ac_1ec_2c_3ic_4$	*Aɛidiw R : ƐDW S : $ac_1ic_2ic_3$
Massette (Roseau)	*Agellu R : GL S : ac_1eC_2u	*Tabuda R : BD S : tac_1uc_2a
Figuier	*Tameɣrust R : MƮRS S : $tac_1ec_2c_3uc_4t$	*Tanqult R : NQL S : $tac_1c_2uc_3t$
Coquelicot	*Taciḥbuḍt R : CḤBḌ S : $tac_1ic_2c_3uc_4$	*Waḥrir R : WḤR S :$c_1ac_2c_3ic_4$
Plante dont les fruits collent a tous ce qu'elles touchent	*Timenṭeḍt R : MNṬḌ S : $tic_1ec_2c_3ec_4$	*Ihinṭeḍ R : ḤNṬḌ S : $ic_1ic_2c_3ec_4$
Variété de la figue	*Tajenğalt R : JNĞL S : $tac_1ec_2c_3ac_4t$	*Tajenjirt R : JNJR S : $tac_1ec_2c_3ic_4t$
La vigne	*Ajgagal R : JGL S : $ac_1c_2ac_3ac_4$	*Tajnant R : JN S : $tac_1c_2ac_3t$
Nigelle	*Ssanuğ R : SNĞ S :$C_1ac_2uc_3$	*Zrareɛ R :ZR□ S :$c_1c_2ac_3ec_4$
Ver	*Tawekka R : WK S : tac_1eC_2a	*Takečča R : KČ S : tac_1eC_2a
Mouton	*Axerfi R : XRF S : $ac_1ec_2c_3i$	*Ikerri R : KR S :ic_1eC_2i
Escargot	*Aɛrus R : ƐRS S : $ac_1c_2uc_3$	*Aɛarus R : ƐRS S : $ac_1ac_2uc_3$
Tortu	*Afekrur R : FKR S : $ac_1ec_2c_3uc_4$	*Ifekker R : FKR S :$ic_1e\,C_2ec_3$

Guêpe	*Areẓ R : RẒ ***S : ac_1ec_2**	*Arẓ***a**ẓ R : RẒ ***S : $ac_1c_2ac_3$***
Chevreau	*Aḥ***uli** ***R :**ḤL** ***S : ac_1uc_2i***	*Iγid R : ΓD S : ic_1ic_2
Variété de figue	*Abakur R : BKR S : $ac_1ac_2uc_3$	*Abukar R : BKR S : $ac_1uc_2ac_3$
Fenouil	*Lbesbas R: (L) BS S : $c_1c_2ec_3c_4ac_5$	*Abesbas R : BS S : $ac_1ec_2c_3ac_4$
Olivier sauvages	*Aḥ***ecca**ḍ ***R :**ḤCḌ ***S : $ac_1eC_2ac_3$***	*Aẓebbuj R : ẒBJ ***S : $ac_1eC_2uc_3$***
Alfa	*Ḥ***lafa** ***R :** Ḥ***LF*** ***S : $c_1c_2ac_2a$***	*Ḥ***lafa** R : Ḥ***LF*** ***S : $c_1c_2ac_3a$***

Annexe 3 :

Corpus : Sahel / Bouzeguène
(Transcription en jaxe)

<?xml version="1.0" encoding="ISO-8859-1" standalone="no" ?>
- <CORPUS>
<CORPUS> <NOTATION USUELLE> : Nekkni zik, ad d-neker deg yiḍ, nṣaf n yiḍ ad nruḥ ta ad teɣɣar i ta, ad d-nagem d talla n wadda mi i d-newwed ad necɣel seksu, mi nfuk seksu-nni, ad nɛedi ad nneneḍ leɣbar mi nfuk leɣbar-nni ad nɛedi ad nerfed iqeṭṭaren tasebḥif; ad nruḥ ɣer cɣel. Nettewqam amarḍil Ad naweḍ a yelli ɣer uzemmur, ad nawi iqcer n uɣrum deg yiciwan n naɣ; ur nettawi ara llɛali-agi i ttawin akka medden tura, wellah ar d tidet a yelli. Ad nawi iqceran-nni n uɣrum deg yiciwan nnteɣ ad naweḍ aken nemwellah d tislatin d lxalat, deg mi ara nali ɣef lǧedra alama n fuk-itt-id deg yixef, mi ara d-nars, aɛeqqa, ad awḍeɣ ar lǧedra ad xezreɣ tazemurt ma ufiɣ aɛeqqa ar teqacuct ad qleɣ, ad t-id-yeqḍeɣ, ḥemlaɣ arrezq a yelli, maci am tura; lǧil n tura. Ad d-nars a yelli tameddit n wass ad aɣ-ineɣ lazz ad nettḍeqir iqceran-nni n uɣrum. “Tecfad yema-m ad tt-ig Rebbi n ṛṛeḥma ; Setti-m ad tt-ig Rebbi n lǧennet. Ula d yema-m tleḥqed”. Ad nettḍegir aɣrum-nni akka ar sdat ad nluqeḍ, ad nettḍegir aɣrum-nni ar sdat ma nufa-d aɛeqqa aquran ad t-nsexḍel s uɣrum-nni. Ur nesɛi reǧwaz ur nesɛi, d aɣrum-nni kan, ad d-naweḍ s lfarḥ d ameqran ad d-naf tabbarbuct am iqeccaḍen ad tt-nečč d tazidant, d tuzyint. Ad nruḥ ma i yefuk uzemmur-nni d tuga, ma tefuk tuga-nni d nqec n tebḥirin, d timegriwin, d inurrar, d tiɛelafin nyezgaren. Ssarwaten madden, deg yiwǧǧiben d tayarza, lxalat d azemmur; irgazen d tiyarziwin n zik d ifellaḥen merra, iɛeqliyen ad mexartayen nneɛma akk d yirden ;merra ad mxartayen. Kul lexir yettzid imir, kul lexir yettzid. Tura d nkkeẓ i nekkẓen lerẓaq ad imnaɛ Rebbi lmumen;d nkkeẓ, d nkkeẓ i nekkẓen larẓaq tura. Ur sineɣ ara ad hedreɣ a tifaryi i yɛedan fell-aneɣ. Alah, alah nniɣ-as a sidi Rebbi ur iɣelli yiṭṭij ar d-nfak, wa ar d-nfak, wa sidi, aRebbi ur iɣeli yiṭṭij-inna ar d-nfak tirni-inna. Nettfaras; ad teẓred diɣ a yewqam-iw asmi wteɣ tagut s aḍu, yemut urgaz-iw deg xemsa wetlatin, d amecṭuḥ; yeǧay-d rebɛa igerdan, tamurt tella nḥend-ik a Rebbi n cekr- ikTenza tbarquqet, tenza tremant, yenza ifelfel, tenza, deg mi ara d-nekker, tamurt n lefni n ddunit newweḍ-itt. Newweḍ Zubga, at aɛbella, newweḍ a yelli Ifarhunen, newweḍ kulci s laḥmul n ukarmus, deg mi ara d-nekkar; hur, hur, hur, ass kamel; hur, hur, hur, ass kamel d tikliwin

ad d-naweḍ, nezenz, ad d-neččar laḥmul-nni d ker; d lebsel, d lebaṭaṭa, ad nuḥ neznuzuy s yidrimen, tamurt tella a yelli nexdem nečča nḥend-ik a Rebbi n cekr-ik tura wellah a Rebbi ar siniɣ di leɛmer feɣaɣ ar lexla ar akka id ufiɣ iman-iw ttɣimiɣ akka deg uxxam. Ad nenɣec, ad nezdem, ad d-nawi aman di lefni n ddunit mi ara čačarent telliwa ar wasif ara nruḥ deg yiḍ, wellah ar daxel n temdiwin id nettaččar tibettiyin, nettruḥu-d s axxam. I iyeɛedan ihi n leqwanen zik. A zik tabarbuct ma teččiḍ-tt aḥeq Rebbi ar xir n miya u miyin d aksum n tura; tabarbuct nzik akka-tt, timɣarin ma ulac ; ma tfuk teyenat; tekarfat n unebdu ad d-tili lbecna, ma tefuk lbecnatabarkant ad d-tili temelalt, ad d-yili ubeluḍ, d azidan lqut, kulec d azidan d awenɛan. Tura, timɣarin arssant alqec, telmezyin tteddunt ɛaryan a llah ibarek d aya i yellan tura, hata win i yellan. D tidet neznuzuy lleft; yettemɣay-d lleft, nettawi ttɛebga n lleft, imir yelluẓ lḥal, ctaqen medden lqut. Ad neččarttɛebgat n lleft art mura ad ad ɣ-d-fken ablluḍ; d tifrac n ubeluḍ ad t-id-nẓed s tesirt ad d-neggar aɣrum; tabarbuct d taẓidant, aɣrum-is d aẓidan, ticki ulac tifrect-nni ntteks-d azegzaw. Tura, azegzaw tura ma tegreḍ-t-id d aɣrum wellah ma tmenaḍ ad d-yekes deg yimi-k, nettedez amaqcur ad t-negar d aɣrum. A Rebbi di tmexluqt-agi taṛwa taḍsa, tariḍ-aɣ d iɛegunen aɛɛi, a yamzuɣ-im. Ikem kan i umi id-hedreɣ akka wanag lami ad d-hedraɣ ma ur ssineɣ lehdur. Aḥeq Rebbi ma sneɣ a yelli lehdur, ḥaca ayen yelan akka sufela. Tenna-as Seɛdiya tariḥant -ad tt-idker Rebbi s lxir, tmeṭut n Lewnis At cilatt- aken ara ad d-awḍen-ken tinni-as : « ayu;ulac lwexda fell-i alama walaɣ-ten beran-d i yiserxuḍen nsen sddaw uxxam n Juhra n tɛeẓugt ». Di lḥara-nni n Juhra n tɛezugt wina ufella, «ayu, a ysetma timaɛzuzin, ulac lwexda fell-i ad wayiɣ bran-d i yisarxuḍen nsen seddaw uxxam n Juhra n tɛezugt ». Ad d-awḍen a yelli, lxallat ad ttgejgijent deg yixxamen, ad tent-zuɣu ; zuɣurn-aɣ amzun d lmal, win weɛan kan ad tsefden s tmeghelt, win weɛan ad tsefden s tɛekazt, ad at-semken s tɛekazt. I yɛedan di lgira-nni i yɛedan, i yɛedan, ad d-naweḍ ayelli taɛzizt-iw aɣ-jemɛen ama ar tejmaɛt, ama ar yiwet n lḥara n ttdakal mera, nettugad. Irgazen a yelli fuken-ten, fiḥel ad am-d-iniɣ, fiḥel, irgazen fuken. Surtu ma ddzn-d a Tefrit, d Buɛwen, d At ɛica. Akken ara ten-walin yergazen tteddun-d wid-ak, ttedun-d lɛeskar-agi, wanag lɛeskar n yiṛumiyen; zik d lmal i ykesen aṛṛebiɛ armi d ass mi i llan akka ixabiten ifuken tafart n yirgazenAh; ayu ass-agi ddan-d aTefrit, d Atɛica, d Buɛewen. I qedren deg yirgazen, i sɛewjen deg yimegra n yirgazen msakit. Aha dayen tura si finin. Ah!di tleqqamt n At aɛli iwsawen, ih ! a lexwayar n yimir, d tirebaɛ mera, d tirebaɛ ; d Arezqi n Welḥağ, Latamen At mecqant, setti-im ad ttig Rebbi n ṛṛeḥma, jedi-im, axxam ahaxxam n At winaten, ad yig Rebbi ččan iqaray n sen; uxxam At sɛada, aken kan ticki nettuqrab akka nettuqrab amɣar-nneɣ ad t-yig Rebbi n ṛṛeḥma, akka, At sɛada ttnusun dina yii-d ass ass-agi ad n-nsen At sɛada deg yiger n tqayed jemaɛ

liman ma neka-tt ula d neknni, ad nemger deg yiḍ ayaxir semeḍ lḥal, ad netta, ad ruḥen At sɛada-nni ad d-feken yiwet di teslatin nsen ad arnuɣ nekki neɣ Tasaɛdit Tamusat, ad nruḥ ad nawi imensi ad nens ai lexla ad nemger, deg yiḍ ad nemgar, deg zal ad nestaɛfu imir arraw-is ! Ah! Tewwi-iyi-d iğadarmiyen acu ara m-id hedreɣ ɣef uraw-iw a yelli, nenuɣ, nenuɣ tewwi-iyi-ten-d Tillult. Ruḥ tura mayella win ara as-yinin diri-tt assa. Ɛecrin n ssnin-agi ur iyi-iluɛa Buxalfa a tarwa n tɛebbuḍt. Haca Ferḥat ; sɛiɣ yiwen kan, wamag Buxalfa tiwwi-t Tillult. Ih! Ay wteɣ tagut s aḍu ɣef uraw-iw yeğğa-ten-id d imecṭaṭaḥmerra, sekreɣ-ten-id s uyenat; s lbiɛw cra. Anda im-nniɣ akka wwḍeɣ aɛi?. Wwḍeɣ Zubga, wwḍeɣ At lɛarbi, wwḍeɣ taddert u, Ifarḥunen, wwḍeɣ abrid Gnnaris ; ak, ak yelan d tamurt merra, merra, akka; Illulen merra di lqern almi d lqern. Mḥaga, d At ɛica, Agrsafen, d yiɣil n Bukyasa, d Tifrit Umalek, iy merra, merra timura-agi merra nenuda-tent-id s lbiɛ w cra. Zzit ad idu, lwarq ad idu, ih ur nesɛi baba-tenaɣ, ur nesɛi yema-tenaɣ, yema ad tt-yig Rebbi n ṛṛeḥma tɛewen-itideg uraw-iw. A! ad nekker, nwet tagut s waḍu, tura i inekr-iyi Buxalfa, a iwwiɣ n daɛwesu. <CORPUS> <TRANSCRIPTION PHONETIQUE>:

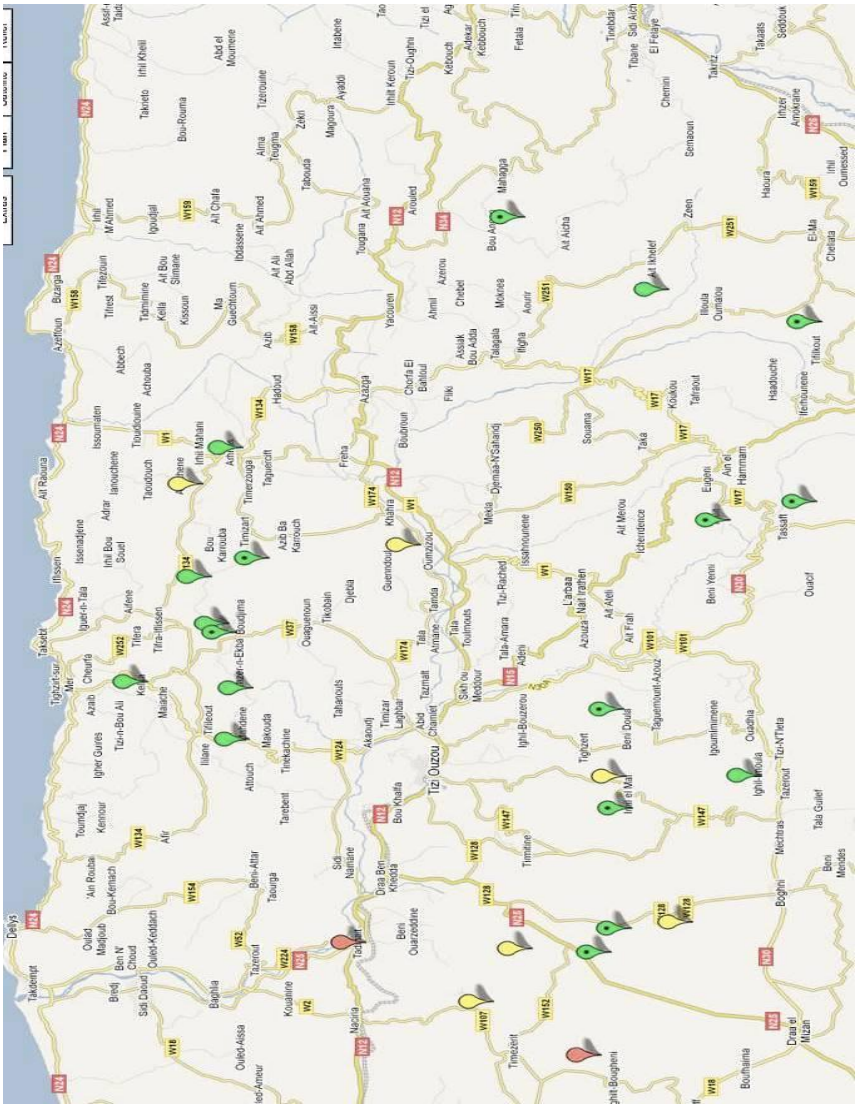
[nəkʷniziχadnəkaʁðəgiðnsːafəgiðanoħeaatsəʁːarieaadnaɣʷemðieajːjapaddːamidneppəðane ʃʁojsəkso:minfoksəkso:nːianʕədianənːnəðjaʁvarminfokjaʁvarnːiaʕədianrfeðiqətarəneasəvħ ieannroħʁərʃʁojnetsəwqamamrðijanawoðajelːːiarozəmːmoranwiiqʃarəppaʁromðəgʃiwannːtaʁ ontsawijarajʕajijagiːtswinakːkaməddənwolːːlahartsiðətsajelːːianawiiqʃrannːnippaʁromðəgʃiwan ntaʁanawəðakənnəmwolahtsislðəjxajaəðəgmaranajiafəldʒðraajamanfokitsidðəgaxəfmaradn arsaʕəqqaadawðaʁaldʒeðraaðxəzraʁeazəmoremaofiʁaʕəqqaareqaʃoʃeaðqəjaʁaeidjəqðaʁħə mjaʁaʀːrəzqajəlːːimatʃameoraaʀːʒʒijntoraadnarsajəlːːieamədiəppasaʁinaʁjazːzaanətsðəgiriqʃər annːniəpwʁromeʃfadjəmːmamatsiɣrəppinəʀːraħmasətsimatsɣrəppiːlːʒənnəeolaðjemːmameəjəħ qədanətsðəgiraʁromnːniakaarzːzaəmanofadaʕəqqaaqoranaənəssəxðəjswəʁromnːniurnəsʕiʁdʒ wazurnəsʕðaʁromnːnikanadnawədsjəfaʀħðaməqranadnafəavːvarvoʃəamiqəʃʃaðənatsnetʃttsaz iðantsozjintanoħmajfokozəmːmornːnitsoɣamaəəfokəoɣanːniðnqəʃetəvħirintsimeɣʀiwinðinora rtsiʕelafinəgəzggarənsarwaəenməddənðəgwəddʒivəntsajarzaijxajaəðazəmːmorirgazəntsijarzi winnziχðifəlːːlaħənmaʀːraiʕəqlijənaðəmxarəajənːnaʕmaakːwəðjirənmaraaðmxarəajenkolijxirj ətsziðimirkolijxirjətszиðəoraðənkːkəzinekːkəznjarzaqəoraursinaʁaraðhəðraʁaəifaʀːʁiigʕəddan fəlːːlanaʁalahalahnːniʁasɑ:siðireppiurijʁəlːːlijiːttiʒardənfakwa:rdənfakwa:siðiɑ:rəppiurixəlːːlijiːttiʒinːnaardənfakəirnijinːnanətsfarasaðatsʒrəððixajəwqamiwasmijəwəaʁəayoəsaðojəmoəwargːgaziwijxəmsawəəjaəinðaməʃtoħjədʒajidʀəvʕaigːwardaneamoreəəlːːlaəamoreəəlːːlɑnħəndikaʀe ppinʃəkrikəənzaəvarqoqəəənzaəerəmantjənzaəjfəjəfəjəənzaðəgmaradnəkːkareəamoreələfnin ddoniənppəðitsnəppəðzovgaaeəʕəvəlːːlaaeəlʕarvinəppəðajəlːːliifarħonənnəppəðkoləʃsjaħmojo karmosðəgmaradnəkːkərhorhorhorasskaməjtsiχliwinadnawəðanəzənzadnatʃarjaħmojnːniðak

arðəjəvsəjðjvatatаanoɦnəznozojsjəðrimənəamoreəəlllaajəlllinxəðəmnəttʃanɦəndikaɾəppinʃək
rikəorawəlllahaaɾəppiarsiniʁðijaʔmarfiʁaʁalaxjaarakkaidofiʁimaniwtsʁimiʁakkagwuxxama
nənʁəʃanəzðəmadnawiamanðiləfninddoniəmaɾatʃatʃarəntəliwaarwasifaranoɦdəgiɖwəlllahar
ðaxejtəməðwinidəntʃattʃareivətsijinntsroɦodsaxxami:igʔədanihiəlaqwanənziχɑ:ziχəavarvoʃ
əmaəəttʃiɖtsaɦəqɾəppiarxirnmi:jawmi:jinðaχsomntoraəavarvoʃənziχakkatsəimʁarinmawla
ʃmaəəfokəəjənatsəχwarfatsonəvðoatsijilvəʃnamaəfokləvʃnaəavarχantatsiliəməlaləaðjíllliov
əloɖðaziðanijqoəkoləʃðaziðanðawənʔaneoraəimʁarinarssantijqəʃəijməzjintsədontʔarjanəlla
hivarəχðajaigəlllaneorahaəawiglan
tsiðətsnəznozojllləfejətsəmʁajdllləfenetsawitsʔəvganllləfeimirjəlozɬhalʃəaqənməddənlqoəad
nətʃartsʔəvganllləfəaremoraaðaʁdəfkənavəloɖtsifraʃovəloɖatsidnəzəðsəəsireadnəgaraʁrome
avarvoʃətsaziðantaʁromisðaziəiʃkiolaʃəifraʃnnintsəksədazəgzawəoraazəgzawəoramaəəɣrət
idðaʁromwəlllahmaəmənaɖaðjəkəsðəgmmiχntsədəzamaqʃoraðəənyarðaʁromaɾəppiðiəmaxj
oqəagiəaɾwaəaɖsaəariɦaʁðiʔəgwonənaʔʔiajaməzoʁimiχəmkanimidhəðraʁakkawanaɣʒʒam
miaddhəðraʁmorsinəʁlahðorɦəqɾəppimasənaʁajəlliləhðorɦaʃaajənjəllankanakkasofəllaəənn
ajassaʔədijaəariɦantatsiðəkərɾəppisjxirəəmətoənləwnisaəʃilatsakənaradawəɖənkanəinasajo
olaʃlwaxdafəlliialamawalaʁəənvərandijsaɾxoɖennəsənsəddawpəxxamnʒohraTʔəʒoɣəðilhara
nninʒohratʔəzoɣəwinaofəllaajoɑ:jəsəəmaəimaʔzozinolaʃlwaxdafəlliarðwajiʁvrandisərxoɖə
nnsənsəddawwəxxamnʒohratʔəzoɣəadawɖənajəlljijəxajaəaðsəgəʒgiʒəntðəgəxxaməna əəntəz
oʁzoʁorənaʁamzonðjmajwinwaʔankanaəsəfðənsəməghələwinwəʔanaəsəfðənsəʔəkwazəəa
ðtssəməχənsəʔəkwazəəigʔədanðilgiranniigʔədanadnawəɖajəlliəaʔzizəiwaʁʒəmʔənamaarəə
ʒmaʔəamaarjiwəəlɦaranətsðakajməɾɾanətsagwaðirgazənajəlljifokəntənfiɦəlamdiniʁfiɦəlirga
zənfokənsortomaddandatəfrieðvoʔwənðaəʔiʃaakkənaraəənwalinjargazəntsdondwiðaχtsdon
djʔəsəkaragiwanaɣijʔəskariɾomijənziχðjmajigχəsənaɾɾəviʔarmiðassmíllanakkaixabieənifok
əneafareəgargazənɑhajo:assagiddandatəfrieðaəʔiʃaðvoʔwəniqədrənðəgargazənisʔew ʒənðəg
məgraɖəgargazənmsaχieahðəajəneorasifininahðieləqamtnaəʔjiiwsawənihaləxwajarəgimirts
irəbaʔməɾɾatsirəbaʔðarəzəqipəlɦadʒlatamənaəməʃqansətsimatsiɣaɾəppinɾɾəɦəmazədimaxxa
mnaəwinaəənaðiɣaɾəppittʃaniqarajnsənwəxxamaəsəʔaðaakənkaneiʃkinətsoqravakkanətsoqr
avamʁarnaʁaəiɣaɾəppinɾɾəɦmaakkaaəsʔaðatsnosənðinajinədassagiaðnəsənaəsʔaðagiɣarətq
ajənðʒəmaʔjimanmanəkkatsojaðnəknianəmɣardəgiɖajaxirsəməðəjɦajaðnətsaaðroɦənaəsʔað
anniadəfkənjiwəəðieəsjaəinnsənaðanərnonəkkinəʁəasəʔðieəamosatsanoɦanawiimənsianən
sðijaxjaanəmɣardəgiɖanəmɣarðəgzajanəseaʔfoimiraheəppijididʒadarmijənaʃoaramdhəðraʁ
ʁafarrawiwajəllinənoʁnənoʁəəppijieənideíllolərоɦeoramajəllawarasjininðiritsassaʔəʃrinləs
najagiorjiloʔavoxajfaaəarwatəʔəboɬhaʃafarɦaesʔiʁjiwənkanwanaɣvoxajfaəəppieəíllolеihaje
weaʁəaɣoəsaɖoʁafarawiwjəʤajinidðiməʃtataɦməɾɾasəkraʁnidsojənaəsjviʔwəʃraandaməni
ʁakkapɖaʁaʔipəɖaʁzovgapəɖaʁaejʔarvipəɖaʁəaddareifarɦonənpəɖaʁavriðgnnarisχaχajəla

ntsmorəəməɾɾamə͡ɾɾaakkaɪllojənmə͡ɾɾaðilqarənajamaðjqarənmɦaɣaðaəʕiʃaaɣwrsafənðjiʁijn voχjasatsifriəomajəχijmə͡ɾɾamə͡ɾɾaəimorajagimə͡ɾɾannoðaəəntsjviʕwəʃraziəaðjədojwarqaəj ədoihornəsʕivavaənaʁornəsʕijəmaənəʁjəmaatsiɣaɾəppinɾɾəɦmaəʕawənijiðəgarrawiwɑadnə karnəwəəəaɣwoəswaðoəoraijnəkərijivoxajfaajppaʁndaʕwəssu]

- <Métadonnées>
<T>traditions et coutumes</T>
<L>Village: Sahel/ Commune: Bouzeguène / Daira: Bouzeguène/ Wilaya: Tizi-ouzou</L>
<D>novembre 2007</D>
<Dial>langue kabyle</Dial>
<Tr>I- Notation usuelle, II- Transcription phonétique</Tr>
<Enreg>méthode semi-directive / caméscope (audiovisuel)</Enreg>
<Loc>Nna ouardia/ sexe : Féminin/ âge : 79 ans/ monolingue</Loc>
<Enq>Karima HABBI</Enq>
<FichVideo href="c:\..." />
<Rem />
</Métadonnées>
</CORPUS>

Annexe 4:


Tizi Ouzou
Azazga
Draa el Mizan
Tigzirt
Boghni
Tizi-N'Tleta
Ouacif
Tadmaït
Tala Guilef
Beni Douala
Tizi-Rached
Tala-Amara
Ain el Hammam
Tassaft
Ain el Kebira
Tizi-n-Bou Ali
Tizi-n-Tleta
Tigzirt-sur-Mer
Takerietz
Delys
Ouled Madjoub
Ben N' Choud
Ouled-Keddach
Bagtiha
Ouled-Aïssa
Bouhalima
Beni Ouazzoûzine
Beni Khelida
Dra'a Ben Khedda
Sidi Naâmane
Naciria
Tadmaït
Tala Amara
Tala Tazmalt
Tala Ouzmou
Tala Toulmoûn
Tala Ouzmou
Tala Ouzmou
Tala Ouzmou

Annexe 5 :

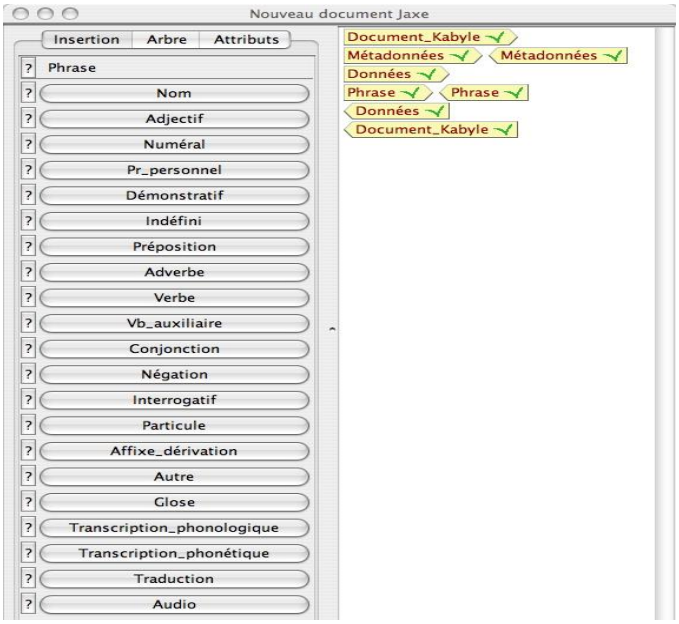
```
<?xml version="1.0" encoding="ISO-8859-1"?><JAXECFG>
<DESCRIPTION>Configuration pour corpus de kabyle -
UMMTO-UNIL</DESCRIPTION>
<RACINE>
<BALISE nom="CORPUS" titre="Référence du corpus"
type="division">
<TEXTE/>
<SOUSBALISE nom="Métadonnées"/>
<SOUSBALISE nom="Données"/>
</BALISE>
</RACINE>
<MENU titre="Métadonnées">
<BALISE nom="Métadonnées" titre="Informations sur ce
corpus" type="division">
<TEXTE/>
<SOUSBALISE nom="T"/>
<SOUSBALISE nom="L"/>
<SOUSBALISE nom="D"/>
<SOUSBALISE nom="Dial"/>
<SOUSBALISE nom="Tr"/>
<SOUSBALISE nom="Enreg"/>
<SOUSBALISE nom="Loc"/>
<SOUSBALISE nom="Enq"/>
<SOUSBALISE nom="FichSon"/>
<SOUSBALISE nom="FichCarte"/>
<SOUSBALISE nom="FichPhoto"/>
<SOUSBALISE nom="FichVideo"/>
<SOUSBALISE nom="Rem"/>
</BALISE>
<BALISE nom="T" titre="Titre" type="string">
<TEXTE/>
</BALISE>
<BALISE nom="L" titre="Lieu" type="string">
<TEXTE/>
</BALISE>
<BALISE nom="D" titre="Date" type="string">
<TEXTE/>
</BALISE>
<BALISE nom="Dial" titre="Dialecte(s)" type="string">
<TEXTE/>
</BALISE>
<BALISE nom="Tr" titre="TypeTranscription"
type="string">
<TEXTE/>
</BALISE>
```

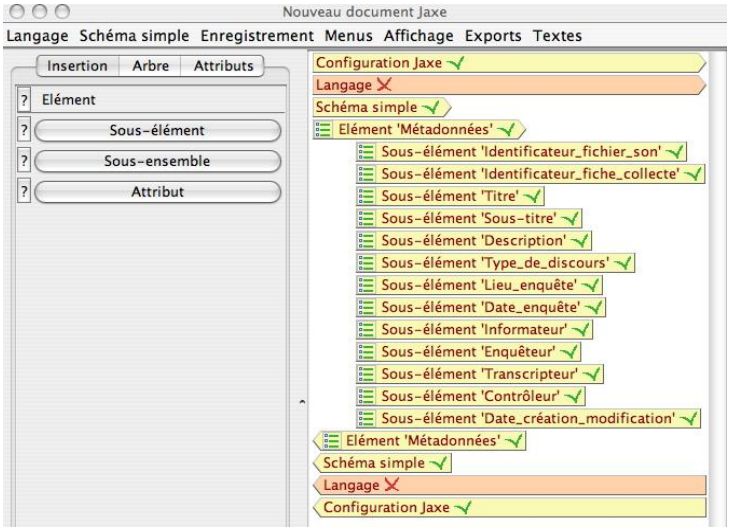
```
<BALISE nom="Enreg" titre="InfosEnregistrement"
type="string">
<TEXTE/>
</BALISE>
<BALISE nom="Loc" titre="Informateur(s)" type="zone">
<TEXTE/>
</BALISE>
<BALISE nom="Enq" titre="EnquÍteur" type="string">
<TEXTE/>
</BALISE>
<BALISE nom="FichSon" titre="Son" type="zone">
<ATTRIBUT nom="href" presence="obligatoire"/>
<TEXTE/>
<ATTRIBUT nom="start" presence="optionelle"/>
<ATTRIBUT nom="stop" presence="optionelle"/>
</BALISE>
<BALISE nom="FichCarte" titre="Carte" type="string">
<ATTRIBUT nom="href" presence="obligatoire"/>
<TEXTE/>
</BALISE>
<BALISE nom="FichPhoto" titre="Photo" type="string">
<ATTRIBUT nom="href" presence="obligatoire"/>
<TEXTE/>
</BALISE>
<BALISE nom="FichVideo" titre="Vidéo" type="string">
<ATTRIBUT nom="href" presence="obligatoire"/>
<TEXTE/>
</BALISE>
<BALISE nom="Rem" titre="Remarques" type="zone">
<TEXTE/>
</BALISE>
</MENU>
<MENU titre="Données">
<BALISE nom="Données" titre="Texte transcrit"
type="division">
<TEXTE/>
<SOUSBALISE nom="Enoncé"/>
<SOUSBALISE nom="Phrase"/>
<SOUSBALISE nom="Mot"/>
<SOUSBALISE nom="Monème"/>
<SOUSBALISE nom="Trad"/>
<SOUSBALISE nom="Phono"/>
<SOUSBALISE nom="Phonet"/>
<SOUSBALISE nom="Graf"/>
<SOUSBALISE nom="ChLg"/>
</BALISE>
<BALISE nom="ChLg" titre="Autre langue" type="string">
<ATTRIBUT nom="code_langue" presence="obligatoire"/>
```

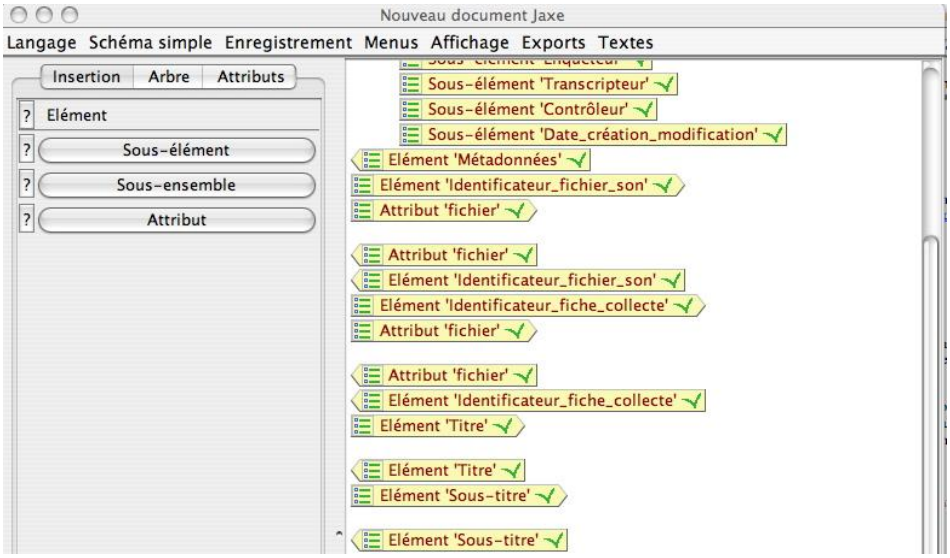
```
<TEXTE/>
</BALISE>
<BALISE nom="Enoncé" titre="Enoncé" type="string">
<ATTRIBUT nom="id" presence="obligatoire"/>
<SOUSBALISE nom="Phrase"/>
<SOUSBALISE nom="Mot"/>
<SOUSBALISE nom="Monème"/>
<SOUSBALISE nom="Trad"/>
<SOUSBALISE nom="Phono"/>
<SOUSBALISE nom="Phonet"/>
<SOUSBALISE nom="Graf"/>
<TEXTE/>
</BALISE>
<BALISE nom="Phrase" titre="Phrase" type="string">
<ATTRIBUT nom="id" presence="obligatoire"/>
<SOUSBALISE nom="Enoncé"/>
<SOUSBALISE nom="Mot"/>
<SOUSBALISE nom="Monème"/>
<SOUSBALISE nom="Trad"/>
<SOUSBALISE nom="Phono"/>
<SOUSBALISE nom="Phonet"/>
<SOUSBALISE nom="Graf"/>
<TEXTE/>
</BALISE>
<BALISE nom="Mot" titre="Mot" type="string">
<TEXTE/>
<SOUSBALISE nom="Monème"/>
<SOUSBALISE nom="Trad"/>
<SOUSBALISE nom="Phono"/>
<SOUSBALISE nom="Phonet"/>
<SOUSBALISE nom="Graf"/>
</BALISE>
<BALISE nom="Monème" titre="Monème" type="string">
<TEXTE/>
<SOUSBALISE nom="Trad"/>
<SOUSBALISE nom="Phono"/>
<SOUSBALISE nom="Phonet"/>
<SOUSBALISE nom="Graf"/>
</BALISE>
<BALISE nom="Trad" titre="Traduction" type="string">
<TEXTE/>
</BALISE>
<BALISE nom="Phono" titre="Phonologie" type="string">
<TEXTE/>
</BALISE>
<BALISE nom="Phonet" titre="Phonétique" type="string">
<TEXTE/>
</BALISE>
```

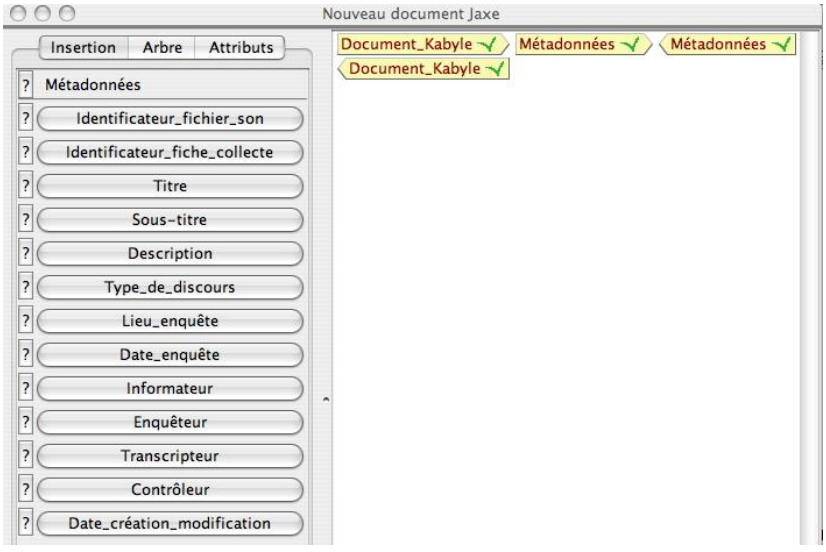
```
<BALISE nom="Graf" titre="Graphie" type="string">
<ATTRIBUT nom="systeme" presence="optionelle"/>
<TEXTE/>
</BALISE> </MENU> </JAXECFG>
```

Annexe 6, 7, 8 et 9 :






Nouveau document Jaxe
Langage Schéma simple Enregistrement Menus Affichage Exports Textes
Insertion Arbre Attributs
? Elément
? Sous-élément
? Sous-ensemble
? Attribut
Sous-élément 'Transcripteur'
Sous-élément 'Contrôleur'
Sous-élément 'Date_création_modification'
Elément 'Métadonnées'
Elément 'Identificateur_fichier_son'
Attribut 'fichier'
Attribut 'fichier'
Elément 'Identificateur_fichier_son'
Elément 'Identificateur_fiche_collecte'
Attribut 'fichier'
Attribut 'fichier'
Elément 'Identificateur_fiche_collecte'
Elément 'Titre'
Elément 'Titre'
Elément 'Sous-titre'
Elément 'Sous-titre'


Nouveau document Jaxe
Insertion Arbre Attributs
? Métadonnées
? Identificateur_fichier_son
? Identificateur_fiche_collecte
? Titre
? Sous-titre
? Description
? Type_de_discours
? Lieu_enquête
? Date_enquête
? Informateur
? Enquêteur
? Transcripteur
? Contrôleur
? Date_création_modification
Document_Kabyle
Métadonnées
Métadonnées
Document_Kabyle

Annexe 10 :

Le corpus ségmenté :

- <Bruit de fond> (00 : 00 s. – 00 : 03 s.)

B- ‖ Ad wen-d-hedrey | ɣ ef ***cinquante-huit*** ? ‖

SPV synt. Prépos.

A- ‖ Ah ? ‖

- <Bruit de fond> (00 : 06 s. – 00 : 08 s.)

A- ‖ **<euh : >** ***alors*** **<euh : >**

‖ Aqlaɣ | di Tesga-Mellul | d ssebt |

Présentatif syntagme nominal syntagme nominal

tnayen-u‿ecrin ɣ uct | ttes‿a ɣ ir r̤r̤be‿ ‖

syntagme nominal syntagme nominal

C- ‖ ɣ ir r̤r̤be‿ ‖

syntagme nominal

B- ‖ Aqlaɣ | di Tesga-Mellul | ttes‿a-wa:crin… |

Présentatif syntagme nominal ?

d ssebt | ttes‿a-wa:crin di ɣ uct |

syntagme nominal syntagme nominal

ssa‿a | attan | d tes‿a **<u: (hésitation)>**

indicateur de thème présentatif syntagme nominal

ɣ ir r̤r̤be‿ ‖

syntagme nominal

- <Bruit de fond> (00 : 26 s. – 31 : 08 s.)

- ‖ Di ***la pression*** n tmanya-uxemsin | yella | lliɣ |

Synt. prépos auxiliaire SPV

di Micli ‖

synt. prépos.

- ‖ Ass-nni | <amar n imjuhad> ‖

Autonome syntagme nominal

- ‖ Ilaq | ad n*regroupi* | ɣ er tudrin ‖
SPV Synt.Prépos.
- ‖ Nekkenni ‖
- ‖ Iṣubb-iyi-d | Dda Sliman | bessif |
SPV expansion référentielle adverbe
si Micli ‖
Synt.Prépos.
- ‖ Nṣubb-d | nebbweḍ | ɣ er ssbiṭar |
SPV SPV synt. Prép.
nufa-d | *l'embuscade* ‖
SPV expansion directe
- ‖ Axaṭar | n‾eṭṭel | di micli ‖
Subordonnant SPV synt. Prép.
yeɣ li-d | ṭṭlam ‖.
SPV expansion référentielle
- ‖ Netta | d *couvre-feu* |
Pronom personnel indépendant Syntagme nominal
amek | teɣ li | *l'embuscade* amezwaru ‖
interrogatif SPV expansion référentielle
- ‖ Neṛğa | ixeddamen | s ukamyun | mi d-ffɣ en |
SPV expansion directe expansion indirecte
Proposition 1
axaṭer | ur nezmir ara | ad n‾eddi ‖
subordonnant Proposition 2
- ‖ n*traversi* | ɣ er lğiha-agi ‖.
SPV synt. Prép.
- ‖ nniɣ -as | ma yella | nufa | *l'embuscade* |
SPV subordonnant auxiliaire SPV expansion directe
ɣ er zdat | ma ‾eddan-d | ad ɣ enɣ en ‖
synt. Prép. subordonnant SPV (proposition1) (prop.2)

- ‖ <nebbwi-ten | daxel> | yeṭṭerḍeq | ukamyun |
SPV adverbe SPV expansion référentielle
neqqim | d ***cca͡b*** | ger-aneɣ |
SPV cordonnant synt. Prép.
en dehors | ***des personnes de <pipa>*** ‖
??????????????????????????????
- ‖ int-as | i ***la supérieure*** | ***i la mère*** ‖
SPV Expansion indirecte Expansion indirecte
- ‖ nnan-as | ğğan-aɣ | weḥd-nneɣ | di ssbiṭar |
SPV SPV synt. Prép. synt. Prép.
neṭṭes | ***en chirurgie*** ‖
SPV synt. Prép.
- ‖ Akken | d ttnac | n deggiḍ |
Adverbe syntagme nominal synt. Prép.
qel͡en | ɣ er tewrirt | s tsita | s i͡ejmiyen ‖
SPV synt. Prép. synt. Prép. synt. prép.

- ‖ Ass-nni | wwten | Wizan n Muḥend Wa͡meṛ ‖
Autonome SPV expansion directe
- ‖ huzzen-tt | deg ufus ‖
SPV synt. Prép.
- ‖ nɣ an-as | taqcict | ɣ ef yiɣ il-is ‖
SPV expansion directe synt. Prép.
- ‖ nɣ an | aɛejmi ‖
SPV expansion directe
- ‖ nɣ an | alews-is | Sa͡id At Ṭṭaleb |
SPV expansion directe expansion directe
dɣ a | wwintt-id | ɣ er ssbiṭar ‖
connecteur (autonome spécifique) SPV synt. Prép.
- ‖ Amek | i s-xedmen ‖

Interrogatif relatif (proposition relative)

- ‖ Ssawlen ‖

SPV

- ‖…- **<Bruit de fond>**….

- ‖ rrfed-itt ǀ imir ǀ si Micli ǀ ʏ er Tizi-Wezzu ‖

SPV adverbe synt. Prép. synt. Prép.

- ✓ // Azekka-nni, / ɛefsen-aneʏ / aɛwin d yiṣurdiyen //
  Adverbe SPV expansion directe
- ✓ // nnan-as / i baba lḥağ, i baba lḥağ Muḥend / kker //
  SPV expansion indirect SPV
- ✓ // yeʏ li / gar tṣeddarin // ;
  SPV synt. Prép.
- ✓ // yenna-as / lukan d lɛibad / i aʏ -d-wwin //
  SPV subordonnant syntagme nominal proposition relative
- ✓ // ur yettsuʏ u ara / s yiṣurdiyen …**la salle d'eau /** d axxam piṭru //;;
  SPV synt. Prép. …. Syntagme nominal
- ✓ // akken ajirikan n waman, acifun ma jajin iʏ min iʏ amen yiṭu i aʏ -ttawin. //
- ✓ // Wwin qbel / deg yimeqqranen // ,
  SPV adverbe synt. Prép.
- ✓ // wwin / Caɛban n Waɛliqa / ad fell-as yeɛfu Rebbi //
  SPV expansion directe
- ✓ // rnan / ɛmara iqa //
  SPV expansion directe
- ✓ // ad ak-rnun Muḥend Waɛ //
  SPV expansion directe
- ✓ // meqqar … Rnan / dadda-k ṣalaḥ, ṣalaḥ At-Aɛli //
  SPV expansion directe
- ✓ // i walbeɛḍ-nneʏ , //
  synt. Prép.
- ✓ // wwin / Yunes At-Saɛdi //

SPV expansion directe

✓ // ula d netta / ur yeqqar ara //

SnV SV

✓ // d baba-s / i yeqqaren / deg-sen / di ɛmara, //

Syntagme nominal proposition relative synt. prép. Synt. prép.

✓ // uɣ alen-d / ɣ ur-i / ur qqareɣ ara, //

SPV Synt.prép. SPV

✓ // daɣ en Ferḥat Sliman akked Muḥend At-Weɛli –Ulḥusin /

Adverbe nominal cordonnant nominal

✓ // ur qqaren ara irkel //

SPV adverbe

✓ // **mais** mbeɛd …kan xemsa n taddart / ufan / xemsa **limumbr**. //

Cordonnant adverbe synt.nominal SPV expansion directe

✓ // Azekka-nni, ad d-nruḥ / ar taddart //

Adverbe SPV synt.prép.

✓ // sukken-aɣ -d / seg Furbiyen, / dinna s-ddaw Uwrir // ,

SPV Synt.prép. déictique synt. Prépos.

✓ // Dda Muḥend-Ḥemmu fell-as yeɛfu Rebbi /

Indicateur de theme

//yeɛwej-as uttbadri //

SPV Expansion référentielle

✓ // deg uzɣ al **hi d-nemlal li-d nhar** / s ukubri-nsen, / s **les para** //

✓ //…lḥara n xali-k Ibrahim … / ḍerfen-d / xemsa-nni n taddart //

Syntagme nominal SPV expansion directe

✓ // ḥebsen-ten / ar lhiḍ / miṭrayin-ten /…..nɣ an-ten //

SPV synt.Prépos. SPV SPV

✓ // iruḥ ciṭ-nni ……//

SPV expansion directe

✓ // **après** weḥd-s yerra-t / **un peu de grace** …..//

Adverbe SPV

✓ // itekka-as / s yiwet / s **la balle** / s aqerru // ,

SPV synt. Prép. synt. Prép. synt. Prép.

✓ // nejbed-as …// // nenṭel-it ……// // tƐedda ddeƐwa //

SPV SPV SPV expansion référentielle

✓ // …….uʏ alen-d **une deuxième fois** //

SPV

✓ // nekkni / nfaq // ,

Indicateur de thème SPV

✓ // ur d-ufin / …… yiwen / yiwen (**répétition**) / di taddart //,

SPV ????????? synt. Prépos.

✓ **[passage d'un autre informateur] / meẓẓi neʏ meqqer //**

✓ // nekk / sƐiʏ tafunast //,

Indicateur de thème SPV expansion directe

✓ // wwiʏ tafunast / ksiʏ -tt //

SPV expansion directe SPV

✓ //…. Deg uxxam / iruḥ lƐesker //;

Synt.prépos. SPV expansion référentielle

✓ // …daʏ nekkini / ma ur iruḥ ara lƐesker //

Subordonnant SPV expansion référentielle

✓ // tenna-ak tmeṭṭut / dʏ a ad tawiḍ tafunast //

SPV expansion référentielle SPV expansion directe

✓ // nekk / ad rreʏ kan / syagi … //

Indicateur de theme SPV adverbe déterminant autonome

✓ // yuʏ al, / arraw-is / bdan / lǧemƐa //

auxiliaire indicateur de thème SPV adverbe

✓ // ass-nnikat d ssebt / dʏ a i neddukkel //

Adverbe syntagme nominal connecteur

✓ // ama d wigad i xeddmen /

Fonctionnel propositionnel SnV relatif prédicatoide

Wigad ur nxeddem ara; //

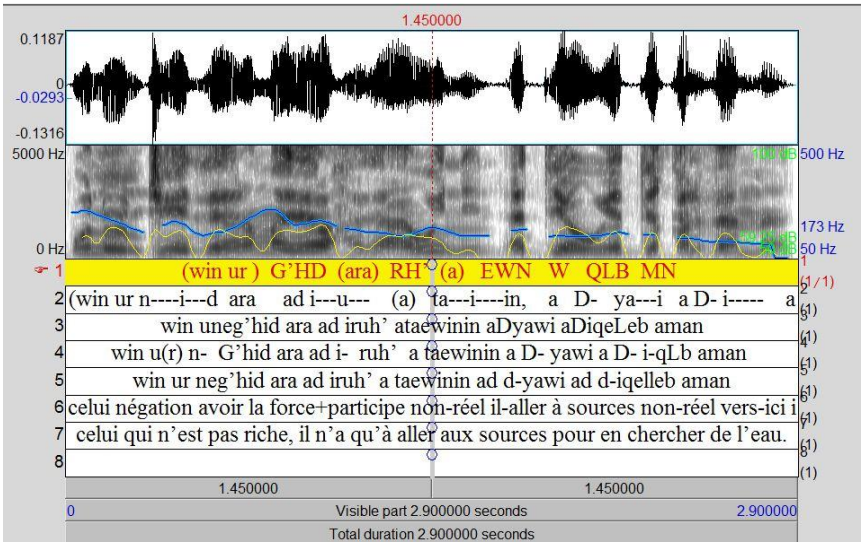
substitut non personnel prédicatoide

- ✓ // yiwen n yiḍ, / tettṣubbu tmeṭṭut //
  Adverbe SPV expansion référentielle
  // tenna-ak / ad awiɣ tafunast //
  SPV SPV expansion directe
- ✓ // niɣ -as / tura ad nens ……. : //
  SPV adverbe SPV

- ✓ // nuqem ttiɛad / ad nemlil / deg yiɣ zer-nni Bu-Sliman //
  SPV expansion directe SPV expansion indirecte

- ✓ // … Xedmen-asen **ratissage** //
  SPV expansion indirecte expansion directe
- ✓ // i ṛemḍan-nni / deg yiḍ, nerǧa / armi d sebɛa //
  Syntagme nominal Synt.prépos. SPV syntagme prédicatoide
- ✓ // xedɛen-aɣ akk //
  SPV adverbe
- ✓ // qqimeɣ -d / ala weḥd-i / di teswiɛt-nni / n Sid-Lḥusin //
  SPV synthème adverbial synt. prépos. Synt.prépos.

Annexe 11 et 12 :


1.450000
0.1187
0
-0.0293
-0.1316
5000 Hz
0 Hz
500 Hz
173 Hz
50 Hz
1 (win ur) G'HD (ara) RH' (a) EWN W QLB MN
2 (win ur n----i---d ara ad i---u--- (a) ta---i----in, a D- ya---i a D- i----- a
3 win uneg'hid ara ad iruh' ataewinin aDyawi aDiqeLeb aman
4 win u(r) n- G'hid ara ad i- ruh' a taewinin a D- yawi a D- i-qLb aman
5 win ur neg'hid ara ad iruh' a taewinin ad d-yawi ad d-iqelleb aman
6 celui négation avoir la force+participe non-réel il-aller à sources non-réel vers-ici i
7 celui qui n'est pas riche, il n'a qu'à aller aux sources pour en chercher de l'eau.
8
1.450000
1.450000
0
Visible part 2.900000 seconds
2.900000
Total duration 2.900000 seconds